Description

USING A PARTIAL METAL LEVEL MASK FOR EARLY TEST RESULTS

BACKGROUND OF INVENTION

[0001] The field of the invention is that of integrated circuits, in particular fabricating circuits with design changes.

[0002] In the course of developing an integrated circuit, the designers make use of extensive computer models that predict the electrical performance of a circuit that has been synthesized, e.g. in a computer program that accepts as input the logic requirements of a system and generates a schematic that performs the functions specified.

[0003] Such programs are very sophisticated, but not infallible.

[0004] During the logic technology qualification cycle, one of the major hurdles needed to overcome is performance limitations that prevent the system specification from being realized. This becomes more critical with advanced technology.

[0005] Numerous process changes and model to hardware corre-

lation are needed to assess the performance specification for the technology. The performance assessment is often made on final product chips with four to six levels of metal. The turn-around-time of this trial and error method to achieve final product performance data is significantly long due to extensive BEOL (Back End Of Line) wiring levels used in the product chip. As we incorporate more device types and move towards System On a Chip (SOC) early readout and correlation of product performance becomes critical. The parametric data does not always match the product performance.

[0006] Product chip data at final test are currently correlated back to the inline parametric data taken during the passage of the wafer through the fab. Parametric data is monitored to provide look ahead and indicate whether the hardware heading to product wafer test is healthy and to predict final chip yield. Parametric monitoring has been performed in this way for multiple generations of integrated circuits, providing standard checks on electrical parameters including the transistor characteristics.

[0007] However, the reference macro designs on which the parametric data is obtained cannot accommodate all the numerous design/process variables that reflect all transis-

tors in the actual product chip environment. Often, the inevitable differences between the parametric transistor design vs. product transistor design can result in conflicting data. For example, it often happens that the parametric transistor results measured inline meet the design manual specification and yet the product chip performance results do not meet the specification set for the chip.

[0008] As performance requirements increase, the margin for error or manufacturing tolerance is reduced, so that a deviation in the performance of an individual device may be magnified by the circuit within which the devices is located. The result may be that the circuit does not meet its specification, even though the device does.

[0009] Due to the complex process of fabricating transistors in various environments, it is possible to have different physical shapes for transistors in different environments. It becomes increasingly difficult to maintain transistor shapes constant across the chip with each scaling technology generation. Systematic technology process solutions (across chip solutions) can only resolve this issue to a certain degree.

[0010] In particular, across chip line width variation may produce transistors with different amounts of drive in different

parts of the chip. This may affect the time when a following device is turned on or off and thus change the circuit timing from its specification. In some cases, there may be logic errors when a signal arrives late or early.

[0011] In addition to the foregoing problems that do not depend on the time to pass through the fab significantly, there are other issues that explicitly deal with the time between identification of a problem and implementation of a solution. Often, a performance limiting feature is only identified on fully finished wafers.

[0012] In order to identify such problems, the chips often need to be delayered down to a lower level where a FIB (focused ion beam) probe technique is used to evaluate the actual transistor behavior in the questionable circuit. These standard techniques simply consume too many resources and have very long turn-around-time.

[0013] For example, the transit time of a wafer through a fab takes several months which the exact turn-around-around time varies depending on specific chip design for Back-End-of-Line build and fab loading, so that it takes a significant amount of time for a test batch of wafers to emerge from the fab so that they can be tested more rigorously than is possible for the in-process

testing. Once a problem is identified (and a possible solution found) it then takes a period of months for wafers embodying the solution to make their way through the fab so that tests can be run to confirm that the proposed solution does or does not work.

[0014] There are several current methods for assessing performance of the product

[0015] 1. Actual product performance at speed –This approach needs a final test, and possibly a packaged module. Further, if a problem is seen it can be difficult to correlate the problem with particular devices.

[0016] 2. The use of test modes (in the testing program) to evaluate parts of the circuit – This approach needs both a final test and also additional circuits on the chip (Silicon area = cost) and design effort.

[0017] 3. Test points within the design – This approach provides direct access to devices, but also needs access to the first metal (in order to locate the test points) and the design functions require significant bench work and design effort.

[0018] 4. Chip delayer and Focus Ion Beam (FIB) cutting + wiring of devices within the chip – This approach also needs for the final level to be completed. In addition, we need to

know which device is causing a problem. Further, since this is a labor-intensive process, there will be only a small sample size, so that the particular components examined may not be the immediate cause of the problem (even if they contribute to the problem in a large statistical sample). This approach often fails to identify a problem that can be fixed when there is an inconclusive test output.

[0019] 5. Correlation of parametric data to product performance – This approach also needs test data from the completed chip. It is also not always successful in correlating the circuit problem with the parametric data.

SUMMARY OF INVENTION

[0020] The invention relates to a method of testing the performance of an integrated circuit in which one or more circuits has a test interconnect arrangement formed in an early layer of interconnect.

[0021] An aspect of the invention is the connection of a subset of transistors in an integrated circuit to form a test structure from that portion of the circuit.

[0022] Another aspect of the invention is the connection of a circuit module for testing.

[0023] Another aspect of the invention is the formation of a test arrangement by a lithographic process.

[0024] Yet another aspect of the invention is the removal of the test arrangement and subsequent formation of a standard interconnect structure.

BRIEF DESCRIPTION OF DRAWINGS

[0025] Figure 1 illustrates the sequence of steps in a fab.

[0026] Figure 2 illustrates in cross section a typical interconnect structure of several layers.

[0027] Figure 3A illustrates an alternative test arrangement.

[0028] Figure 3B illustrates an alternative test arrangement using two levels of test interconnect.

[0029] Figure 4 is a plan view of a test chip and neighboring chips.

DETAILED DESCRIPTION

[0030] Figure 1 illustrates in condensed form the progress of a wafer through an integrated circuit facility (fab).

[0031] Box 10 represents steps referred to as the front end of the line, in which various operations are performed on the semiconductor substrate. For example, implants to form N- and P- wells in the substrate, threshold implants, isolation structures, such as shallow trench isolation, etc.

[0032] Box 20 represents the first level of interconnection (sometimes referred to as interconnect) that forms the

electrical circuit by connecting transistors and other circuit components, referred to as "metal 1". Other levels of interconnection are conventionally referred to as "metal 2", etc. The standard metal in the past has been aluminum, which is now being supplemented or replaced by copper. The invention does not depend on the choice of material used for the interconnection.

[0033]   The first level of metal begins the section of the process referred to as the "back end of the line", or BEOL. Some workers in the field also refer to the middle of the line, meaning the contact studs connecting the transistors and first level metal which in-process testing can take place.

[0034]   Referring back to Figure 1, box 25 refers to measuring the parameters of various monitoring structures. Conventionally, integrated circuit wafers have the circuits separated by a space, referred to as the kerf, which is removed during the process of sawing the wafer to separate the circuits before packaging. It is customary to construct test transistors, metal structures, etc in the kerf space. These test structures have contacts suitable to be probed by test equipment. The test structures are tested after various steps to monitor the quality of the various process steps.

[0035]   For example, test transistors may be measured to check

their turn-on voltage, current capacity, etc. In some cases, an error will be found and the particular wafer can be pulled out of the line and reworked by stripping off the defective layer and doing it over. For example, there may have been a misalignment in the second level of metal that causes many chips on the wafer to have a short circuit. Instead of scrapping the wafer, it may be economically feasible to strip off that level of metal and deposit it again.

[0036] It will be evident that the sooner a defective process step is detected, the better. Not only will there be less work and risk in removing a single level than in removing several, there will be fewer defective wafers made when the defect is systematic, rather than an error that applies to a single wafer. For example, an alignment error by the operator may affect only a single wafer. An error in the flow rate of a chemical will affect all the wafers passing through the particular machine having that error.

[0037] Boxes 30, 40 and 50 denote the formation of metal levels in a process having several layers of metal interconnections.

[0038] Box 60 denotes the final product test, typically after the wafer has been separated into chips and packaged. It is

usually not possible to salvage chips that are identified with defects at this stage, and the only thing that can be done is to change the process and scrap the defective chips.

[0039] Arrow 110 in the Figure indicates the standard feedback path in the past, in which the final product test is analyzed to identify which step in the process is faulty and future wafers are processed with a different value of the relevant parameter.

[0040] On the right of the Figure, Arrow 120 illustrates the feedback path according to the invention. In the illustrative embodiment, the test structure comprising transistors, connections, etc. is formed with contacts and interconnects in the first level of metal. If a flaw or a parameter out of specification is detected it is corrected. If correction is not feasible, there will be no need to waste time on that wafer by performing later steps in the process.

[0041] The arrow on the lower right side indicates that in a typical fab the time spent between metal 1 and final test is about 5 weeks and in many fabs it is more.

[0042] Figure 2 shows in cross section a typical integrated circuit. On the bottom, layer 200 represents the semiconductor substrate and a row of rectangles that represent transis-

tors. Vertical lines 202 represent contacts to the sources and drains of the transistors. Contacts to the gates of the transistors are not shown, for simplicity. A layer of dielectric insulates the transistors from one another and from the interconnection members. The deposition process of forming a pattern in photoresist, depositing a layer of metal in the pattern (or of putting down a layer of metal and removing the unwanted portions) will be referred to as a lithographic process in the claims.

[0043] Layer 210 represents the first layer of metal, which extends into and out of the plane of the paper to connect various circuit elements, not only the transistors shown, but also resistors, capacitors, diodes etc.

[0044] Layer 220 represents the second level of metal, having vertical members, referred to as vias, that connect to the metal 1 layer and horizontal members that form various connections as required by the circuit.

[0045] Layer 230, 240, 250 and 260 represent metal 3, metal 4, metal 5 and metal 6, respectively. Layer 270 represents a pad for making contact to the packaging. Pads are conventionally much larger than the lower levels because the packaging contacts can not be aligned as accurately as the lower levels.

[0046] Figure 3A shows a cross section of a test interconnect structure according to the invention, in which layer 312 represents a layer in the metal 1 level above the wafer, but is set up to define a test circuit, which may be as small as a single transistor or as large as a multiplier or other large-scale circuit module. The interconnect components, whether metal 1, and/or metal 2 will be referred to as a test interconnect arrangement.

[0047] Figure 3B show a similar arrangement to that of Figure 3A, but in which the test structure uses two levels, denoted 310 and 320. This arrangement would be used for a more complex test circuit that requires two levels to be laid out. The invention is not restricted to two level and various combinations of higher and/or lower levels may be used. Levels 1 and 2 will be referred to as being close to the substrate, for convenience in the claims. Those skilled in the art will appreciate that early detection of flaws in tran-sistor formation is advantageous.

[0048] In carrying out a test program according to the invention, a separate mask is used for one or more test chips that uses the transistors that have been fabricated as part of the circuit, not special test transistors. The test mask con-nects some of the transistors in the circuit to form a test

structure.

[0049] In some cases, the test structure (containing a subset of transistors in the circuit) may be identical with a portion of the final circuit (a circuit module), as in a situation where a circuit module depends critically on transistor or other parameters in order to work properly. In other cases, where the final circuit uses higher levels of metal, the test structure may achieve the same result using one or two lower levels of metal.

[0050] In some other cases, the test structure may be assembled out of transistors used for other purposes – e.g. the circuit contains an assortment of AND circuits, OR circuits, etc (referred to as logic unit circuits), that are connected by the test mask to form a test circuit that is not part of the final design but provides useful information on the quality of the process.

[0051] The test process will usually comprise applying a known input (a test vector) to a test subset of transistors and comparing the output with the connect output vector.

[0052] In operation, a special mask having the test circuit(s) can be used for a subset of chips on a wafer. Steppers often have storage areas to hold masks used in various levels of wafers or for various wafers. The regular mask for metal

level 1, say, would be used for most chips on a wafer and the test mask(s) would be used for those chips that have the test circuits.

[0053] Workers in the field known that wafers have "sweet spots" that consistently have higher than average yields of good chips. Conversely, areas that consistently have lower than average yield will be referred to as sensitive locations. Preferably, the location of test arrangements will concentrate on sensitive areas that strongly depend on variations in process parameters. The number and location of test locations will depend on a tradeoff between time spent in testing, lost chips and the benefit in chips saved by more or less closely monitoring the process.

[0054] In a typical stepper, a process referred to as step and repeat comprises the steps of moving the stepper to a chip location, exposing the pattern of a chip and repeating the process by moving to the next chip. If the chips are small, the stepper can expose more than one chip at a time.

[0055] Figure 4 illustrates an example in which brackets 410 and 420 define an area of four chips that are exposed in a single step of the stepper, one of which is a test chip. Two kerf strips 422 are located at the top and bottom of the Figure. In alternative layouts, there may be kerfs between

each chip. In this example, the test mask will have the regular layout for chips 2, 3 and 4, with the test layout for chip 1. Those skilled in the art will readily be able to arrange for other test masks for different steppers and chip sizes. If the chips happen to be large, then each chip in Figure 4 may be a single stepper location.

[0056] The ability to evaluate transistors or circuits inside a performance path gives a great advantage in terms of understanding the limiter issue and getting a resolution for a performance problem. The ability to get this information by using a special metal mask(s) enables readout at lower level metal, and minimizes a significant amount of time in the learning and evaluation process. This saving is particularly significant during early product evaluation cycle.

[0057] This allows quicker turn-around-time to isolate performance limiters and make proper process changes to resolve the issue. This technique avoids the lengthy time required to do a full product evaluation with each and every new process improvement. It enables a significant amount of statistical data gathering as every chip can be measured. This means earlier and improved correlation to parametric data, which will eventually be used to monitor the process.

[0058] Evaluating circuit performance limiters at transistor level using a special wire level mask to gain learning, results in significant time and cost saving. Currently standard end of the process product chip characterization is often followed by chip delayer (stripping off upper layers) and FIB probe to evaluate performance limiter at FET level. This method imposes a significantly higher cost and time consumption compared to the present invention.

[0059] The invention does not depend on the type of semiconductor (Silicon, silicon-germanium, gallium arsenide, etc.) or transistor, (bipolar, FET, etc.), type of interconnect (damascene or not, aluminum or copper).

[0060] In an alternate mode, the test interconnect layer(s) may be stripped and replaced with the standard pattern of that level, e.g. if the test wiring is on M1, the wafer has the M1 level on that chip stripped and redone. If more convenient, the entire wafer could be stripped of M1 and redone. Whether this alternative is employed will depend on the relative cost of rework and the value of what is salvaged. If the only benefit is salvaging the test chip, rework may not be worthwhile. If experience has shown that rework of the defect revealed by the test can salvage 1/3 or 1/2 of the chips on the wafer, the decision is more likely

to go in favor of rework.

[0061] While the invention has been described in terms of a single preferred embodiment, those skilled in the art will recognize that the invention can be practiced in various versions within the spirit and scope of the following claims.